



CDF SAM Deployment Status

Doug Benjamin
Duke University
(for the CDF Data Handling Group)

CDF's definition of SAM Deployment

- Raw data logging only into SAM database schema
- Production Farm writes only to SAM
- User analysis of new data (collision/ simulated) available SAM only

(from my GDM talk 30-Aug-05)

Beyond Deployment

- Common Root ntuple from production data
 - CDF has two widely used (~ 80% of the collaboration) ntuple formats. Should be considered as production-level datasets.

SAM access to production data (Users)

- **Major Success!!!**
 - Been using SAM v7 client and db servers
 - since 15 September. > 300 TB collision data.
 - Ntupleing of data proceeding well.
 - (> 26 TB of common ntuples produced)
- Applied the strategy of minimizing the impact on the users - to keep their productivity up
 - User's scripts required small changes
 - Fully tested the examples ahead general usage.
 - Good documentation

Deployment Status

- Completed
 - Production farm
 - User access to production data via SAM
- Incomplete
 - Raw data logging
 - Calibration ntuple creation executable still uses DFC schema
 - Monitor backup cron job (Predator) to verify it does not move metadata; 4-6 weeks of monitoring (raw data too valuable to lose) (started on 1-Nov-05)
 - MC upload - testing has begun

Common Ntuples and SAM (further issues)

- Root Ntuples are how most CDF users access the data.
 - Two major ntuples types (Standard Ntuple and Top Ntuple ~ 80% CDF users)
- These Ntuples should be in the data handling system
- SAM use cases
 - Batch - Users macros loop over files (like production files - `diskcache_i`) - *want minimal changes to users' macros*
 - Requires SAM interface in Root (based on C++ API)
 - Interactive tests to ensure successful batch jobs (mimic batch tests interactively as a test)
 - Interactive data exploration
 - Requires SAM tool to download files to desktop

Appendix - Slides with further details

Current Status - Raw Data Logging

- Raw Data logging still requires three computers until CDF Consumer Server Logger (CSL) is upgraded ('06 - '07) - CDF online responsibility.
- Dehong Zhang has made the system more robust
 - Cross mounted NFS mounted disk used to transfer metadata (ascii files) between sgi machines and Linux SAM machine
 - Sam meta data writing has automatic retry in case of errors (V7 client).
- Raw Data metadata still logged into Data File Catalogue (DFC) **and** SAM DB schemas
 - Job to create ntuples for Calibrations - still uses DFC - Currently offline operations managers (Aidan Robson and Bernd Steltzer) volunteered to fix situation. (Extraordinary effort that we are very grateful for!!!)

Current Status - Production Farm

- Production Farm - SAM based
 - Can process > 22 M events/day
 - Issues:
 - Running v6 of SAM client/ DB server
 - Working w/ SAMGrid team to test v6 SAM client/ v7 DB server.
 - SAMGrid team has identified two changes to client API
 - Migration of Prod. Farm to SAM v7 client proceeding slowly -
 - Production group responsibility moving from the Taiwan group (Suen H & Tsan L) to the UNM (Elena Vataga)
 - Several weeks of running v7 SAM on test farm is required before using v7 client on the production farm. - ***to maintain robust farm performance.***
 - With smaller farm need to process > 50 M events to ensure success.

Current Status - MC Data Upload

- Monte Carlo generated offsite
 - Generation 5 (v5.3.* offline software - 2004) - uses DFC tools to save MC data on tape (and DFC schema)
 - DFC schema is copied to SAM schema daily
 - The Physics groups set the schedule for Gen 5 MC
 - Generation 6 (v6.1.2> offline software 2005 >) will use SAM tools (SAM_upload)
 - Tools developed/ maintained by Armando Fella (and others from Italy) (SAM_upload)
- CDF MC production group starting tests of SAM upload tools ahead of large scale MC production

Current Status - MC Upload (2)

- Currently CDF private Autodest server.
- The official SAM autodest server needed CDF features added
 - Steve White, Randolph Herber and Valeria Bartsch worked implementing the CDF specific needs into the SAM version
 - Testing has just begun
 - My estimate: at least 6 weeks before CDF is using official SAM autodest server.
 - Due to testing and script modifications (Sam_upload and Production farm scripts)